

WORD-FORMATION MODELING

Liudmyla Chumak (Zhytomyr, Ukraine)

New words are appearing continuously in various domains. The new word identification is an indispensable task to speech recognition, optical character recognition, handwriting recognition, machine translation, spelling correction. The task of language modeling is to predict the next word in a sequence of words [1]. Adding a prefix or suffix, combining or blending words all create new forms.

Word-formation means that a lexeme originates from (an)other one(s) [2]. Word-formation modeling is used to search for appropriate alternatives that may be found in morphological variants or in case of usage phrase structures. The constituents of a phrase structure are phrase types or word types. A phrase type specifies the structure of a phrase, which can be used as a model element name. Word types are the elements of a phrase type. And at the same time word types act as placeholders for particular words.

A word type consists of a distinct word class (a noun or a verb, an adjective, an adverb, a numeral, a pronoun, etc.) and its inflection (that can be specialized as case, number, tense, gender, mood, person, etc.). Independently from their corresponding word class, particular uninflected words are called lexemes (e.g. the verb “*work*”). Inflected words are called word forms (e.g. past participle “*worked*”). Word forms are assigned to the corresponding word types. To this point, word forms represent lexemes of a particular word type. The word types are characterized by the features word class and inflection.

Traditionally, a word can be divided into the minimal linguistic units that bear meanings or grammatical functions (i.e. morphemes). There are four criteria of what a unit takes to be a morpheme: it should have a meaning or function, recur in other words with a related meaning (e.g. *un-* > *unhappy*), and be involved in a pattern of interchange (e.g. *-est* in *deepest* can be substituted with another morpheme such as, *-er*).

Morphemes are further categorized into lexical morphemes (e.g. *-full*, *-ness*, etc.) and grammatical morphemes or inflections (e.g. *-ed*, *-s*). So, it's necessary to

accentuate that the development of English inflectional morphology differs from that of derivational morphology.

The study of morphology has been approached by two complementary approaches: analytic and synthetic [3: 12-20]. The analytic approach is concerned with morpheme identification or breaking a word down into its meaningful components, as, for example, in the word “*notebooks*”: *note-book-s*. The synthetic approach, on the other hand, is concerned with productivity of morphological structure or bringing the smallest pieces (morphemes) together to form words.

Various models have been proposed to account for how morphological units are encoded and decoded. The qualitative research of new words implicates the data collection methods (observations, interviews, visual images, and documents) and data analysis methods (that need the following steps: identification, classification, interpreting, describing, conclusion) [4: 7-13].

References

1. Jelinek F., Merialdo B., Roukos S., Strauss M. A dynamic language model for speech recognition. In Proc. Wkshp on Speech and Natural Language, 1991. – pp. 293–295, Pacific Grove, California. ACL.
2. Becker J. Towards increased comparability of conceptual models – enforcing naming conventions through domain thesauri and linguistic grammars. In [Conceptual Modeling - ER 2009: 28th International Conference on Conceptual Modeling, Gramado, Brazil, November 9-12, 2009, Proceedings \(Lecture Notes ... Applications, incl. Internet/Web, and HCI\)](#) by Alberto H. F. Laender, Silvana Castano, Umeshwar Dayal and Fabio Casati (26 Oct 2009), 2009. – pp. 378-390.
3. Aronoff M., Fudeman K. What Is Morphology? (Fundamentals of Linguistics). Wiley-Blackwell, 2005. – 257p.
4. Wray A., Trott K., Bloomer A. Projects in linguistics: a practical guide to researching language. London: Arnold / NY: Oxford. 1998. – 303 p.